RESOURCES
Molecular Ecology Resources (2013) doi: 10.1111/1755-0998.12139

Comparison of molecular species identification for North Sea calanoid copepods (Crustacea) using proteome fingerprints and DNA sequences

S. LAAKMANN,* G. GERDTS,† R. ERLER,† T. KNEBELSBERGER,* P. MARTÍNEZ ARBIZU* and M.J. RAUPACH*

**Senckenberg Research Institute, German Center for Marine Biodiversity Research (DZMB), Südstrand 44, 26382 Wilhelmshaven, Germany, †Alfred Wegener Institute for Polar and Marine Research, Biologische Anstalt Helgoland, Kurpromenade 201, 27498 Helgoland, Germany*

Abstract

Calanoid copepods play an important role in the pelagic ecosystem making them subject to various taxonomic and ecological studies, as well as indicators for detecting changes in the marine habitat. For all these investigations, valid identification, mainly of sibling and cryptic species as well as early life history stages, represents a central issue. In this study, we compare species identification methods for pelagic calanoid copepod species from the North Sea and adjacent regions in a total of 333 specimens. Morphologically identified specimens were analysed on the basis of nucleotide sequences (i.e. partial mitochondrial cytochrome *c* oxidase subunit I (COI) and complete 18S rDNA) and on proteome fingerprints using the technology of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). On all three molecular approaches, all specimens were classified to species level indicated by low intraspecific and high interspecific variability. Sequence divergences in both markers revealed a second *Pseudocalanus* species for the southern North Sea identified as *Pseudocalanus moultoni* by COI sequence comparisons to GenBank. Proteome fingerprints were valid for species clusters irrespective of high intraspecific variability, including significant differences between early developmental stages and adults. There was no effect of sampling region or time; thus, trophic effect, when analysing the whole organisms, was observed in species-specific protein mass spectra, underlining the power of this tool in the application on metazoan species identification. Because of less sample preparation steps, we recommend proteomic fingerprinting using the MALDI-TOF MS as an alternative or supplementary approach for rapid, cost-effective species identification.

Keywords: 18S rDNA, COI, MALDI-TOF MS, proteomic fingerprinting, species identification, zooplankton

Received 26 March 2013; revision received 31 May 2013; accepted 4 June 2013

Introduction

Among the zooplankton, calanoid copepods show high biological and ecological diversity. This makes them one of the most studied marine taxonomic groups focusing on diversity, morphology, taxonomy, phylogeny, distribution, life-cycle strategies, feeding behaviour and adaptation to various environmental conditions (e.g. Mauchline 1998 and references therein; Bradford-Grieve *et al.* 2010; Blanco-Bercial *et al.* 2011; Saiz & Calbet 2011). Due to their high abundances and biomasses, extensive worldwide distributions across seas and oceans, and because they link various trophic levels, calanoid copepods represent a major component of the marine plankton community, food web and pelagic ecosystem (e.g. Longhurst 1985; Fransz *et al.* 1991). They link primary production and higher trophic levels (e.g. Mauchline 1998; Dam & Lopes 2003; Calbet *et al.* 2007) and play a key role in the dynamics of economically important fish stocks (e.g. Möllmann *et al.* 2003). The diversity of calanoid copepods is likely to be mainly regulated by environmental temperature, hydrodynamics, stratification, seasonal variability and water masses making their patterns of occurrence as an environmental indicator to assess changes in the marine habitat (Beaugrand *et al.* 2002). Therefore, they are often used as indices and indicators for detecting changes and shifts in marine ecosystems (e.g. Beaugrand 2004; Edwards & Richardson 2004), whereby a basic pattern of changes in the pelagic ecosystem can be rendered by the sum of the abundance of copepod populations (Greve *et al.* 2004). The analysis of

Correspondence: Silke Laakmann, Fax: +49 4421 9475111;
E-mail: slaakmann@senckenberg.de

long-term time series data (e.g. Continuous Plankton Recorder data) showed contrasting results by analysing biomass and diversity (e.g. Beaugrand 2004). For example, the interpretation of categories like 'total copepods' was shown to be not straightforward, as this group can encompass a large number of species, thus cannot detect possible changes in the community structure (Beaugrand 2004).

For all these studies, valid species identification represents a central issue to identify sibling and cryptic species as well as different life history stages for getting insights into population structures, abundances, diversity, recruitment, secondary production and long-term changes in the pelagic environment. Because morphological identification by microscopy can be challenging, time-consuming, and requires a strong taxonomic background, the demand on alternative species identification methods represents a central issue for analysing organisms on species level as well as for detecting qualitative (diversity) and quantitative (abundances, biomass) changes in marine community and habitats. These alternative methods should fulfill the requirements of being rapid, comparable low priced, simple to be performed, cost-effective and accurate.

In the last decades, many kinds of molecular techniques for unambiguous species detection, identification and discrimination were developed. Most of them are promising or already proved to be reliable approaches, which facilitate or overcome taxonomic difficulties in the identification of species and life history stages. As an example for DNA sequence analysis, the DNA barcoding approach based on the analysis of a fragment of the mitochondrial cytochrome *c* oxidase subunit I gene (COI) has been proved as reliable tool for the identification for animals (e.g. Hebert *et al.* 2003; Bucklin *et al.* 2010a,b, 2011; Radulovici *et al.* 2010). For copepods, increasing numbers of studies applied COI for the identification of species and to elucidate intraspecific patterns of variability (e.g. Lee 2000; Bucklin *et al.* 2010a,b; Chen & Hare 2011; Laakmann *et al.* 2012). In addition to this approach, other gene fragments like mitochondrial 16S rDNA (e.g. Bucklin *et al.* 1995; Rocha-Olivares *et al.* 2001; Goetze 2003, 2010; Caudill & Bucklin 2004), nuclear ITS2 (Rocha-Olivares *et al.* 2001; Goetze 2003; Laakmann *et al.* 2012) or even conserved 18S rDNA (Bucklin *et al.* 2003; Goetze 2003) are frequently used for the discrimination of copepod species. To exclude sequencing, species-specific PCR assays were developed for a rapid identification (e.g. Hill *et al.* 2001; Grabbert *et al.* 2010), as well as PCRs with adjacent restriction fragment length polymorphism (RFLP) were analysed to differentiate even between different life history stages (e.g. Lindeque *et al.* 1999). Next to sequencing, other molecular techniques like DNA hybridization (e.g. Kiesling *et al.* 2002) or single-gene zooplankton community analysis (Machida *et al.* 2009) were conducted. Beside DNA investigations, techniques on analysing proteome fingerprints are promising. For example, the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) is well established for identifying microorganisms (i.e. bacteria, viruses, fungus and spores) in diagnostic bacteriology (e.g. Holland *et al.* 1996; Haag *et al.* 1998; Fenselau & Demirev 2001) and may represent a useful method for a valid species identification of zooplankton taxa. In this approach, the sample (whole cells, molecules, extracted peptides or proteins) cocrystallizes with a matrix solution on a target plate. A pulsed laser causes desorption of the sample–matrix, followed by the ionization of the sample. In a strong electric field, the ions are accelerated and drift along a vacuum tube. Based on the time of flight, the different masses of the single molecules are represented as spectra. For metazoans, some MALDI-TOF MS pilot studies were conducted using species-specific proteome profiles for a successful and rapid species identification (e.g. Mazzeo *et al.* 2008; Feltens *et al.* 2010; Kaufmann *et al.* 2011; Volta *et al.* 2012), including for three closely related freshwater copepod species (Riccardi *et al.* 2012).

In our study, we focus on calanoid copepods of the North Sea where they show high abundance and biomass, representing the most important portion of the zooplankton and dominating the smaller part of the zooplankton community for most of the year (e.g. Hickel 1975; Fransz *et al.* 1991). In the German Bight, calanoid copepod fauna is dominated by small populations, mainly by the herbivorous to omnivorous calanoids *Paracalanus* spp., *Pseudocalanus* spp., *Temora longicornis*, *Acartia* spp., *Centropages typicus* and *Centropages hamatus* (e.g. Greve *et al.* 2004). For most of the species, late copepodite- and adult-stages specimens can be identified to species level. In contrast to that, for some particular species determination to species level and thus analysis of population structure (abundance, biomass) is difficult due to the co-occurrence of congeneric species (i.e. *Acartia*, *Pseudocalanus*, *Paracalanus*, *Calanus*). In addition, the morphological species discrimination of nauplii or early copepodite stages represents a time-consuming challenging task using microscopy and is particularly impossible because of lacking diagnostic species characters. Because these neritic and estuarine species do not only play an important role in the North Sea but have a widespread occurrence across the Northern Hemisphere as well as some are cosmopolitans (e.g. Razouls *et al.* 2005-2012), rapid, accurate and cost-effective identification methods are requested and important in a broader context.

The aim of this study is to test the combination of morphological and molecular methodologies for accurate taxonomic species identification by comparing different

identification methods for the North Sea calanoid copepod community. For this, species are identified based on morphological characters, followed by a comprehensive analysis using partial mitochondrial COI, complete nuclear 18S rDNA and proteome profiles using the MALDI-TOF MS technology. Identification results and application parameters will be discussed.

Material and methods

Sampling

Copepods were sampled at different years (2010–2012), seasons and regions in the North Sea, mainly in the southern part, as well as from the shore in Jade Bay (Wilhelmshaven, Germany) and Weser Estuary (Table 1, Fig. 1). To cover intraspecific variability, some specimens were taken from western Baltic Sea and Skagerrak (Gullmarsfjord, Sweden) (Fig. 1). Additional *Acartia tonsa* specimens were analysed from running cultures originating from Kiel Bight (Baltic Sea). Sampling was performed by horizontal sea surface hols (max. depth 3–4 m) using a Calcofi Net (300 and 500 μm mesh size; Hydrobios) and Apstein net (150 μm mesh size; Hydrobios). Immediately after the catch, whole zooplankton samples were fixed in absolute ethanol. Copepod specimens were sorted from bulk samples and identified to species level using microscopy. For further analyses, individuals were isolated in absolute ethanol. Only well-preserved specimens without contaminations on the exoskeleton were used for the analyses.

Molecular genetic analyses

Genomic DNA of the total 137 specimens (adult females and males) of 13 calanoid and one harpacticoid copepod species was extracted using the QIAGEN DNeasy® blood and tissue kit (QIAGEN) following the manufacturer's protocol with an overnight lysis (see Table 1). PCR amplifications were accomplished by *illustra* PuReTaq *Ready-To-Go PCR* Beads (GE Healthcare) using 4 μL of DNA templates in 25-μL reaction volumes. COI amplification and sequencing were performed using the primer pair LCO1490 and HCO2198 (Folmer *et al.* 1994) at annealing temperature of 45 °C for 38–40 cycles. PCR failures were not species-specific, and in these cases, two other primer pair combinations (LCO1490 together with Nancy (Simon *et al.* 1994) or Cop-COI-2189 (Bucklin *et al.* 2010b)) were applied as well as lower annealing temperatures (minimum 42 °C). The complete 18S rDNA was amplified using the primer pair 18A1 mod and 1800 mod (Raupach *et al.* 2009) at an annealing temperature of 50 °C for 36–38 cycles. For sequencing, additionally the primers F1 (5′-AGCAGCCGCGGTAATTCCAGCT-3′), CF2 (5′-GAAACTTAAAGGAATTGACGGAA-3′), CR1 (5′- CCTTCCGTCAATTCCTTTAAGT-3′) and R2 (5′-AGCTGGAATTACCGCGGCTGCT-3′) (this study) were used. PCR products were purified using QIAquick® PCR Purification Kit (QIAGEN). Both, PCR products and purified PCR products were checked on an agarose gel (1%) with GelRed (0.1%). Strands were sequenced using the BigDye™ terminator chemistry and an ABI3730XL automated sequencer (Macrogen, Amsterdam). Sequences were assembled, edited and checked for reading frames using the software GENEIOUS version 5.4.5 created by Biomatters (available from http://www.geneious.com/). Using BLAST (Altschul *et al.* 1990), sequences were compared with those available in GenBank. All new sequences have been deposited in GenBank (see Table 1). Multiple alignments of COI and 18S sequences were performed in MUSCLE version 3.8.1 (Edgar 2004) using default settings and are provided as supporting information data S1 and S2, respectively.

A COI fragment of 657 bp (minimum sequence length 625 bp) from 136 calanoid specimens was analysed by neighbour-joining analysis based on K80 model (Kimura 2-parameter (K2P): equal base frequencies, one transition rate and one transversion rate; Kimura 1980) and 10 000 bootstrap replicates using the software MEGA version 5.05 (Tamura *et al.* 2011). The harpacticoid copepod *Euterpina acutifrons* (JX995145) was chosen as outgroup taxon. Pairwise genetic distances were calculated on K80 model with the same software.

For the 1844 bp alignment of 18S rDNA comprising 43 calanoid specimens, maximum likelihood analysis was performed using RAxML-VI-HPC (Stamatakis 2006) with the GTRGAMMA nucleotide substitution model and the generation of 10 000 bootstrap replicates. The harpacticoid *Tigriopus fulvus* (EU370430) was defined as outgroup taxa. Ungapped minimum and maximum lengths of the fragment were 1737 and 1772 bp. Pairwise genetic distances were calculated with the GTRGAMMA nucleotide substitution model using RAxML-VI-HPC (Stamatakis 2006). For both COI and 18S rDNA data, species clusters based on pairwise genetic distances were tested by analysis of similarity (ANOSIM) with 999 permutations using the software PRIMER6 version 6.1.6 (Clarke & Gorley 2006). For visualizing genetic distances as a heatmap, Excel data on average-specific differences were saved as text file and then opened in FRAMEWORK version 1.2b (Wirth 2009).

Protein mass fingerprinting analysis (MALDI-TOF MS)

The MALDI-TOF MS technology was tested for 197 specimens of 11 calanoid species (Table 1, Fig. 1). For the abundant species, different developmental stages were

Table 1 Investigated species and specimens including sampling information, developmental stage, numbers and Accession Numbers in GenBank

Sampling information			Molecular genetics			MALDI-TOF MS Analysis
				GenBank Accession Numbers		
Latitude/Longitude		Date	Stage and number of specimens	COI	18S rDNA	Stage and number of specimens
Acartia bifilosa						
	53.5085 N; 8.1567 E	June 2010	CVIF(7)	JX995248-54	JX995289-91	
Acartia clausi						
	55.3500 N; 0.1500 W	Aug 2010	CVIF(3)	JX995237-39		
	54.3310 N; 5.6700 E	June 2010	CVIF(4)	JX995240-43	JX995282	
	54.6700 N; 5.6678 E	June 2011	CVIF(2)	JX995244-45	JX995283	
	58.8170 N; 2.6865 E	Aug 2011	CVIF(1), CVIM(1)	JX995246-47	JX995284	
	54.1863 N; 7.9000 E	May 2012				N(1),CI-II(4),CIII(2),CVIF(3)
	54.1863 N; 7.9000 E	Aug 2010				CVIF(6)
	54.5000 N; 7.9997 E	June 2010				CVIF(6)
Acartia tonsa						
	54.3540 N; 10.1673 E	Culture	CVIF(5)	JX995255-59	JX995285-86	
	54.9868 N; 7.9723 E	June 2011	CVIF(5)	JX995260-64		
	53.5085 N; 8.1567 E	July 2011	CVIF(5)	JX995265-69	JX995287	
	54.7958 N; 13.1227 E	Aug 2011	CVIF(1)	JX995270	JX995288	CII-IV(4),CV(7)
Anomalocera patersoni						
	54.1408 N; 7.5984 E	May 2010	CVIM(5)	JX995193-97	JX995305	
	54.1717 N; 7.9047 E	May 2010	CVIM(5)	JX995198-202		CIV(2),CV(3),CVIF(4),CVIM(5)
	58.7837 N; 2.4195 E	Aug 2011	CVIM(5)	JX995203-07	JX995306	
	55.3472 N; 0.0883 W	Aug 2011	CVIF(1), CVIM(4)	JX995208-12	JX995307	CVIF(4),CVIM(4)
	57.8825 N; 0.8313 W	Aug 2011	CVIF(1), CVIM(1)	JX995213-14		
Calanus helgolandicus						
	54.1863 N; 7.9000 E	July/Aug 2010	CVIF(9)	JX995223-31	JX995315-16	CV(5),CVIF(6)
	58.2612 N; 11.4510 E	Aug 2011	CVIF(5)	JX995232-36	JX995317-18	CVIF(6)
	54.1863 N; 7.9000 E	May 2012				CIV(1),CVIF(4)
Centropages hamatus						
	53.6700 N; 5.6700 E	June 2010	CVIF(5)	JX995174-78	JX995292	CVIF(6)
	54.9998 N; 7.6663 E	June 2010	CVIF(1)	JX995179	JX995293	
	54.1863 N; 7.9000 E	May 2011	CVIM(2)	JX995180-81	JX995294	
	54.7560 N; 10.1002 E	Aug 2011	CVIF(1)	JX995182	JX995295	
	54.1863 N; 7.9000 E	May 2011				CVIF(6)
	54.1863 N; 7.9000 E	May 2012				N(1),CI-II(4),CVIF(3)
Centropages typicus						
	54.1863 N; 7.9000 E	Aug 2010	CVIF(2)	JX995165-66	JX995296	CVIF(6),CVIM(6)

Table 1 (Continued)

Sampling information		Molecular genetics			MALDI-TOF MS Analysis
			GenBank Accession Numbers		
Latitude/Longitude	Date	Stage and number of specimens	COI	18S rDNA	Stage and number of specimens
54.5662 N; 7.1468 E	June 2011	CVIF(1),CVIM(1)	JX995167-68		CVIF(5),CVIM(4)
55.3472 N; 0.0883 W	Aug 2011	CVIF(1)	JX995169	JX995297	
57.8825 N; 0.8313 E	Aug 2011	CVIF(1),CVIM(1)	JX995170-71		
58.7837 N; 2.4195 E	Aug 2011	CVIF(2)	JX995172-73	JX995298	
Eurytemora affinis					
53.5085 N; 8.1567 E	April 2011	CVIF(5)	JX995183-87	JX995299-301	CVIF(6)
53.5155 N; 8.5536 E	Aug 2011	CVIF(5)	JX995188-92		CVIF(6)
Isias clavipes					
56.5685 N; 5.3317 E	July 2011	CVIF(4),CVIM(1)	JX995160-64	JX995302-04	CVIF(3)
Paracalanus parvus					
54.5000 N; 7.9997 E	June 2010	CVIF(4)	JX995215-18	JX995311-12	CVIF(2)
58.2612 N; 11.4510 E	Aug 2011	CVIF(4)	JX995219-22	JX995313-14	
Pseudocalanus elongatus					
55.000 N; 7.6700 E	June 2010	CVIF(1)	JX995271	JX995319	CVIF(6)
54.1700 N; 7.7700 E	May 2010	CVIF(2)	JX995272-73	JX995320	CVIF(6)
58.8200 N; 2.6900 E	Aug 2011	CVIF(3)	JX995274-76	JX995321	
54.1863 N; 7.9000 E	May 2012				N(3),CI-II(7),CVIF(4)
Pseudocalanus moultoni					
53.0000 N; 3.0000 E	June 2010	CVIF(5)	JX995277-81	JX995322-24	
Temora longicornis					
53.7200 N; 8.2700 E	May 2010	CVIF(1)	JX995146		
53.0000 N; 3.000 E	June 2010	CVIF(2)	JX995147-48		
54.6700 N; 5.0000 E	June 2010	CVIF(2)	JX995149-50	JX995308	CVIF(6)
55.0000 N; 7.6700 E	June 2010	CVIF(1)	JX995151	JX995309	
53.6700 N; 5.6700 E	June 2010	CVIF(2)	JX995152		
53.5085 N; 8.1567 E	June 2011	CVIF(7)	JX995153-59	JX995310	CVIF(6)
54.1863 N; 7.9000 E	May 2012				N(4),CI(4),CII(2),CIII(2),CIV(3) CV(3),CVIF(2),CVIM(4)
Total analysed specimens		137			197

N, nauplii, C, Copepodite stages, F, female, M, male.

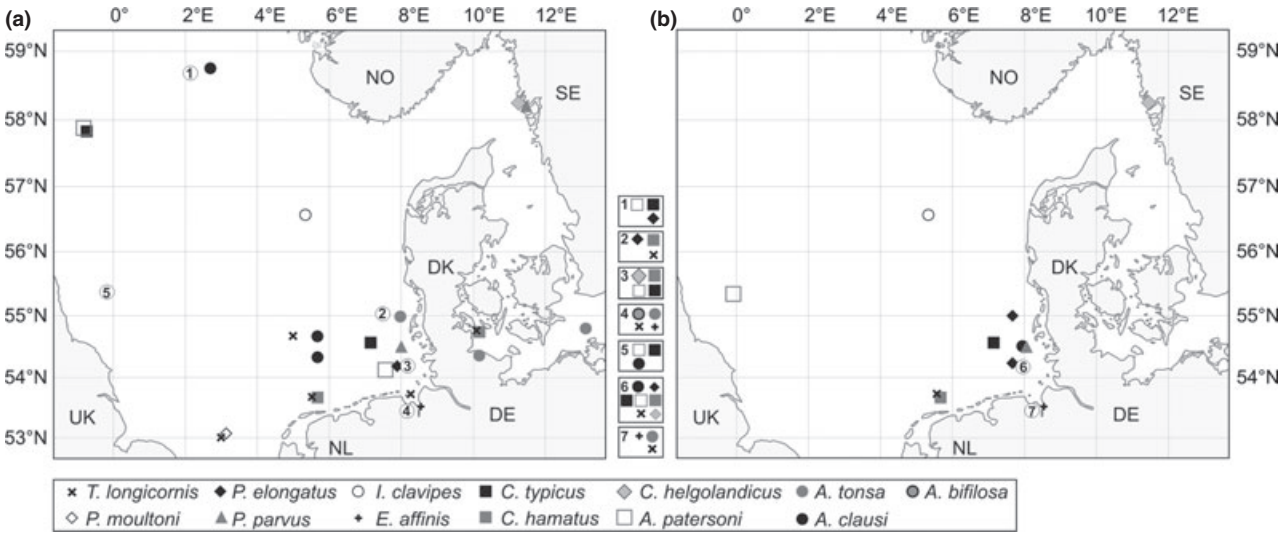

Fig. 1 Sampling localities of specimens for (a) DNA investigations and (b) proteomic fingerprinting (MALDI-TOF MS). Localities where multiple species were sampled are marked with a number, and species list is given in a legend. Countries are abbreviated by country code (DE, Germany, DK. Denmark, NL, The Netherlands, NO, Norway; SE, Sweden and UK, United Kingdom).

analysed comprising copepodite stages I–VI, females and males. Additionally, unidentified naupliar stages were analysed. Because whole specimens were analysed, the influence of trophic and feeding effects was examined by analysing specimens from different stations (and thus geographic regions) in the North Sea as well as from different seasons and years. Single specimens were placed into 5–10 μL of matrix, containing alpha-cyano-4-hydroxycinnamic acid (HCCA) as a saturated solution in 50% acetonitrile and 2.5% trifluoroacetic acid. After an incubation of 10 min at room temperature and dark conditions (Riccardi *et al.* 2012), 1.2 μL of the extract was spotted in three replicates on the target plate and allowed to evaporate at room temperature for several minutes. Each spot was measured three times resulting in nine replicate measurements per specimen. Analyses were performed with the compact linear-mode benchtop microflex LT System (Bruker Daltonics) at a laser frequency of 60 Hz. To create one spectrum, 240 laser shots were generated at fixed optical laser energy and a pulse of 3 ns. Calibration was performed using a Bacterial Test Standard (Bruker Daltonics) containing a protein extract of *Escherichia coli* DH5alpha. The mass spectra were analysed between 2000 and 20 000 Dalton (Da). The analyses of the mass spectra were performed within the software FLEXANALYSIS (version 3.3; Bruker Daltonics) by smoothing using the Savitzky Golay smoothing filter (Savitzky & Golay 1964) with two mass-to-charge-ratio (m/z) and 10 cycles. Baseline subtraction was performed using a Top-Hat filter. Mass list was found on the basis of following settings: centroid peak detection algorithm, 1.5 signal to noise threshold, 300 minimum/maximum number of peaks, one m/z peak width and 90% peak height.

Mass lists were exported as peak tables to the software Excel (S3). The peak values of relative intensities were classified into ascending intervals with 3-, 5- and 10-Da bin sizes. Cluster analysis, analysis of similarity (ANOSIM) for testing significant differences between variables (and data sets), and similarity percentages (SIMPER) for testing which variable is mainly responsible for observed differences were performed using the software PRIMER6 version 6.1.6; Clarke & Gorley 2006). Single-linkage cluster analysis was performed based on a Bray–Curtis similarity matrix of the average values of the samples (maximum nine replicates per specimen). Data sets of the different Da categories (3-, 5- and 10-Da) were compared by ANOSIM with 999 permutations, as well as the 5-Da data set for differences between the species clusters. The differences between protein mass profiles of the different species, stations and stages were additionally tested by Kruskal–Wallis test as well as Mann–Whitney-test using the software GRAPHPAD PRISM (version 6; Windows Demo). To detect strongest mass signals, peak tables were analysed by SIMPER analysis. As already described for the sequence data, a similarity matrix of proteome fingerprints was visualized as heatmap.

Results and discussion

Calanoid copepod biodiversity of the North Sea

In total, 12 calanoid copepod species were successfully identified based on morphological characters: *Temora longicornis*, *Acartia clausi*, *Centropages hamatus*, *C. typicus*, *Pseudocalanus elongatus*, *Paracalanus parvus*, *Anomalocera patersoni* and *Calanus helgolandicus* occurred frequently in

the samples; *Acartia tonsa*, *Acartia bifilosa* and *Eurytemora affinis* were abundant in Jade Bay, the latter species as well in the Weser estuary; *Isias clavipes* was found at only one station (56.57′N, 5.34′E). In contrast to that, both sequence analyses revealed 13 monophyletic groups (Figs 2 and 3), uncovering a second *Pseudocalanus* species, emphasizing as *Pseudocalanus moultoni* by the COI sequence comparison to published sequences in GenBank. This species was sampled at one station in the southern North Sea (53.00′N; 3.00′E). Because interspecific morphological divergences are weak in *Pseudocalanus* species (Frost 1989), the sequence divergences in both mitochondrial and nuclear gene fragments indicated different species. This detection of sibling and cryptic species by analysing DNA sequence data was already demonstrated in several studies (by e.g. Lee 2000; Hill *et al.* 2001; Bucklin & Frost 2009). Originally known from the NW Atlantic (e.g. Bucklin *et al.* 1998), *P. moultoni* specimens were recently identified in North European waters by COI sequencing (Aarbakke *et al.* 2011) and are reported here from the southern North Sea as well. However, due to the sympatric occurrence and the challenging taxonomy of the *Pseudocalanus* species (e.g. Frost 1989; Bucklin *et al.* 1998; Markhaseva *et al.* 2012), it remains unclear whether *P. moultoni* is a recent alien invader into European waters, or whether it has traditionally been overlooked and confused with co-occurring morphological similar congeners (Aarbakke *et al.* 2011) such as *Pseudocalanus elongatus* in southern North Sea. The successful discrimination of co-occurring sibling *Pseudocalanus* species by molecular tools was also demonstrated in other regions (i.e. Georges Bank, Bucklin *et al.* 1998).

The dominant species in the North Sea and adjacent seas were typical for the regions as defined by Beaugrand *et al.* (2002) for Atlantic westerly winds and coastal biomes (i.e. *Para-Pseudocalanus* spp., *C. typicus*, *C. helgolandicus*, *Acartia* spp.) and neritic regions (i.e. *T. longicornis*, *C. hamatus*, *A. patersoni*, *I. clavipes*). *E. affinis* is a typical estuarine species, with a wide distribution range across the Northern Hemisphere (e.g. Winkler *et al.* 2008, 2011). Most of these species show broad distribution ranges across the Northern Hemisphere, and some species also occur in Southern Hemisphere waters (e.g. Razouls *et al.* 2005-2012). In this context, the effort of the three investigated molecular identification methodologies may find application in various pelagic systems and is not restricted to the North Sea and/or adjacent waters.

Species identification based on nucleotide sequences

Species discrimination based on COI and 18S rDNA nucleotide sequences was unambiguous and revealed 13 species (Figs 2 and 3). For pairwise genetic distances (COI and 18S rDNA), ANOSIM revealed a Global R of 1 ($P = 0.001$) for the different species clusters, thus revealing highly significant differences between the species.

Cytochrome *c* oxidase subunit I analysis revealed 13 species with species cluster supported by bootstrap values of 99% (Fig. 2). Genetic variability can be overestimated if pseudogenes are present (Bucklin *et al.* 2000), but this is not a significant issue for the investigated North Sea calanoids as no stop codons were detected in the COI sequences. Interspecific distances ranged from 17.6% to 35.8%, with lower values for congeneric species (i.e. *Acartia*, *Centropages*, *Pseudocalanus*; 21.5% ± 1.0%) than species from different genera (27.9% ± 2.7%; Fig. 2, Table 2). On intraspecific level, variation ranged from 0% to 5.9% (Fig. 2, Table 2). Hence, the barcoding gap was given with at least 11.7% difference between intraspecific and interspecific pairwise genetic K2P distance, leading to a species identification success rate both for distant and closely related species of 100% (Fig. 2). Inter and, in general, intraspecific COI variability was comparable to those of other studies (e.g. Bucklin *et al.* 2003) as well as high haplotype diversity, which was already found for other calanoids across European Seas (e.g. Castellani *et al.* 2012). High intraspecific variations were detected in *A. patersoni*, *P. elongatus* and *C. helgolandicus* with 0–5.87, 0–4.24 and 0–3.21%, respectively (Fig. 2, Table 2), comparable to those of geographically distant populations in other calanoid species (e.g. Laakmann *et al.* 2012). Comparatively low intraspecific variability was found for *A. clausi*, originating from a broad sampling range in the North Sea, as well as for *A. tonsa* sampled in the North and Baltic Sea and for *P. parvus* from the North Sea and Skagerrak (Fig. 2). For *C. hamatus*, only one specimen was analysed from the Baltic Sea. This specimen differed from the North Sea individuals by 1.8–2.0%, a comparatively high divergence in relation to the low intraspecific variation among North Sea individuals (0–0.2%). By visualizing all pairwise genetic distances as heatmap, COI showed a high and unambiguously efficient resolution on species level (Fig. 4a).

Analysis of the whole 18S rDNA fragments revealed all 13 species with bootstrap values ≥ 94% (Fig. 3). No variability was found on the intraspecific level, already demonstrated for other calanoids (Bucklin *et al.* 2003) as consequence of a concerted evolution of rDNAs (e.g. Hillis *et al.* 1991; Eickbush & Eickbush 2007). The lowest and highest interspecific GTRGAMMA distances were 0.5% and 81.4% thus, the gap between intra and interspecific distances was less pronounced compared with the COI analysis (Table 2). Exceptionally high divergence was observed in the genus *Acartia* (belonging to the superfamily Centropagoidea), both within genus and to other calanoids. This high divergence led to gaps in the

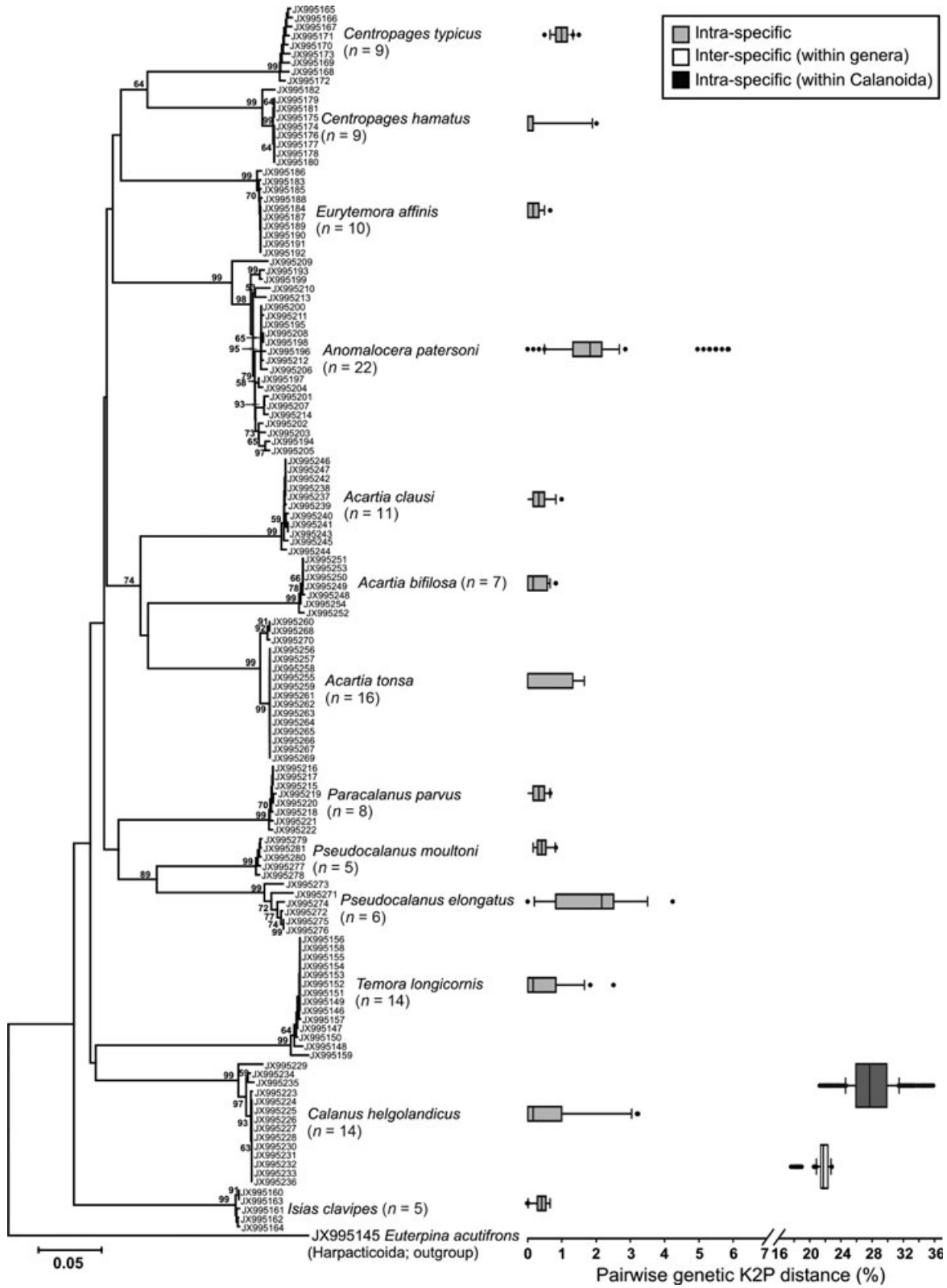

Fig. 2 Neighbour-joining analysis of the 657 bp cytochrome *c* oxidase subunit I (COI) fragment based on K80 model with 10 000 bootstrap replicates. Numbers on branches represent bootstrap values higher than 50%. Species-specific intraspecific and interspecific pairwise genetic distances are represented as box plots (box comprises 50% and whiskers 90% of the data set while dots depict outliers).

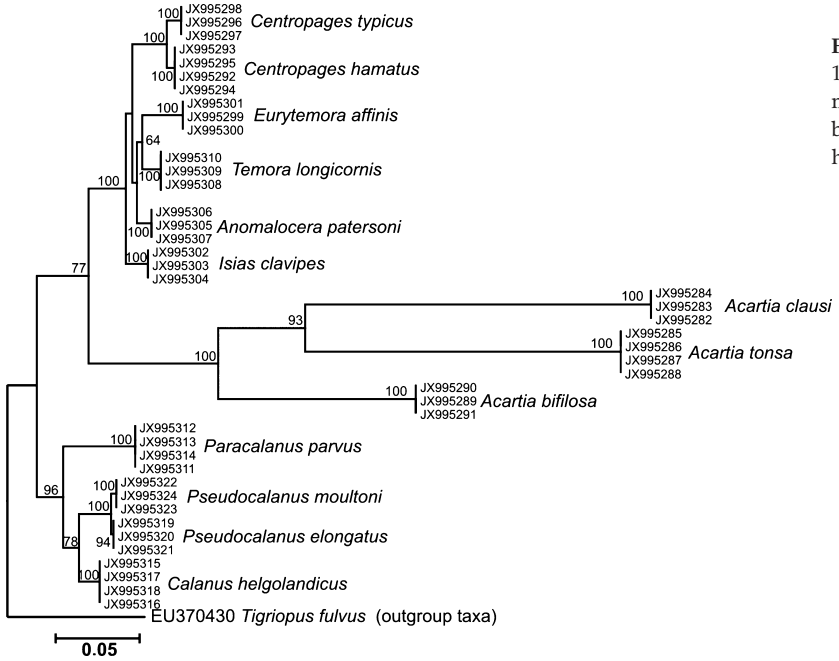

Fig. 3 Maximum likelihood analysis of 1844 bp 18S rDNA fragment based on GTRGAMMA model with 10 000 bootstrap replicates. Numbers on branches represent bootstrap values higher than 50%.

alignment as well as high interspecific distances ranging from 29.7% to 81.4% (Table 2). By aligning these sequences with those of other *Acartia* species obtained from GenBank, a similar pattern was observed, highlighting within this genus an uncommon divergence, which is in need of further morphological and molecular taxonomic studies. The other calanoid species differed by 0.5–14.0% (Table 2). Next to species clusters, 18S rDNA analysis provided information on higher taxonomic levels with lowest interspecific distances between congeners (i.e. *Centropages* 1.3% and *Pseudocalanus* 0.5%, Table 2). Regarding the taxonomic level of superfamilies, the different families and genera of the Centropagoidea (Diaptomoidea) (i.e. *Centropages*, *Isias*, *Temora*, *Eurytemora* and *Anomalocera*) clustered together supported by high bootstrap value (100%) and differed from one another by 2.2–5.4%. Those of the superfamily Calanoidea (Megacalanoidea) (i.e. *Calanus*, *Paracalanus*) differed by 6.7% (Table 2). Higher values were found between superfamilies with 9.8–14% between Centropagoidea and Calanoidea, 10.6–12.9% between Centropagoidea and Clausocalanoidea (i.e. *Pseudocalanus* congeners), and comparatively lower values of 3.4–7.3% between Calanoidea and Clausocalanoidea (Table 2). The variability between species of one superfamily was in accordance with those observed among calanoids (Blanco-Bercial *et al.* 2011). Pairwise genetic distances, visualized as heatmap, demonstrated the identification of the three *Acartia* species but no similarity within this genus, while the other species belonging to the Centropagoidea, Calanoidea and Clausocalanoidea illustrated close relationships, respectively (Table 3, Fig. 4b). Excluding highly divergent *Acartia* from this analysis, the individual clusters of the other calanoid species were more pronounced, except for low diversification of only 0.5% between closely related *Pseudocalanus* congeners (Fig. 4c).

Species identification based on proteomic fingerprinting

The analysis of the total 197 individuals, comprising various developmental stages, revealed reproducible mass spectra, adequate for the classification of the 11 different species. The ANOSIM revealed highly significant differences of mass spectra between the species with global Rs of 0.940, 0.972 and 0.956 and *P* values of 0.001, 0.001 and 0.001 for the 3-, 5- and 10-Da categories, respectively. Therefore, the data set on 5-Da categories was chosen as the best classification for species discrimination, and the matrix is provided as supporting information data (S4). Regarding bulked relative intensities of each analysed specimen with more than 2% contribution using SIMPER analysis, highest peak intensities were recorded between 2000 and 9000 Da, similar to those found for other metazoans (e.g. Feltens *et al.* 2010). The cluster analysis showed species-specific signal patterns for all the 11 species with lower values on intraspecific than on interspecific level, even for congeners (Fig. 5). Except for *P. parvus* and *Isias clavipes* (because of too low sample size), all species differed significantly from one another based on ANOSIM (Table 3) as well as by Kruskal–Wallis test ($P > 0.0001$). The specificity of species proteome

Table 2 Pairwise genetic distances for cytochrome *c* oxidase subunit I (COI) based on K80 model (upper diagonal, italics) and 18S rDNA based on GTRGAMMA model (lower diagonal, regular)

	Ac	At	Ab	Cham	Ct	Ea	Ic	Ap	Tl	Pp	Ch	Pe	Pm
Ac	*0.3 ± 0.3*	*21.4 ± 0.2*	*25.2 ± 0.3*	*28.8 ± 0.2*	*29.1 ± 0.4*	*25.3 ± 0.2*	*33.1 ± 0.2*	*26.7 ± 0.5*	*35.0 ± 0.4*	*28.1 ± 0.3*	*26.7 ± 0.6*	*27.7 ± 0.4*	*24.9 ± 0.3*
At	81.4	*0.5 ± 0.5*	*22.1 ± 0.2*	*27.2 ± 0.2*	*30.7 ± 0.4*	*25.6 ± 0.3*	*32.7 ± 0.2*	*25.2 ± 0.3*	*31.3 ± 0.4*	*27.0 ± 0.3*	*28.8 ± 0.5*	*25.1 ± 0.6*	*21.7 ± 0.3*
Ab	68.2	54.8	*0.3 ± 0.3*	*26.9 ± 0.2*	*26.7 ± 0.3*	*26.0 ± 0.2*	*29.5 ± 0.4*	*28.5 ± 0.6*	*30.2 ± 0.5*	*32.7 ± 0.2*	*28.0 ± 0.3*	*28.0 ± 0.6*	*27.1 ± 0.3*
Cham	56.8	59.1	32.6	*0.8 ± 0.7*	*22.0 ± 0.6*	*24.1 ± 0.3*	*26.0 ± 0.4*	*24.6 ± 0.6*	*30.0 ± 0.3*	*26.4 ± 0.3*	*29.0 ± 0.2*	*31.4 ± 0.5*	*27.7 ± 0.2*
Ct	55.9	59.3	31.7	1.3	*1.0 ± 0.2*	*23.3 ± 0.3*	*28.4 ± 0.4*	*26.1 ± 0.7*	*30.6 ± 0.5*	*29.8 ± 0.3*	*28.2 ± 0.4*	*28.8 ± 0.7*	*28.0 ± 0.2*
Ea	56.4	57.6	31.1	5.2	5.4	*0.2 ± 0.2*	*27.4 ± 0.2*	*23.7 ± 0.4*	*33.1 ± 0.3*	*26.6 ± 0.2*	*28.8 ± 0.3*	*26.8 ± 0.6*	*25.6 ± 0.3*
Ic	54.7	58.8	32.0	4.4	4.6	4.5	*0.4 ± 0.2*	*27.8 ± 0.7*	*30.1 ± 0.3*	*32.7 ± 0.2*	*29.6 ± 0.5*	*32.5 ± 0.3*	*31.4 ± 0.2*
Ap	54.5	56.1	30.4	3.5	4.2	3.7	2.8	*1.9 ± 1.3*	*30.3 ± 0.6*	*25.9 ± 0.4*	*27.0 ± 0.5*	*27.3 ± 0.5*	*24.9 ± 0.6*
Tl	56.2	56.6	29.7	3.9	4.5	3.6	3.4	2.2	*0.4 ± 0.6*	*32.7 ± 0.4*	*29.8 ± 0.4*	*30.5 ± 0.6*	*30.3 ± 0.4*
Pp	67.1	62.8	36.6	13.1	12.9	14.0	12.4	12.2	12.8	*0.3 ± 0.2*	*25.4 ± 0.2*	*23.8 ± 0.4*	*24.9 ± 0.2*
Ch	61.7	63.0	35.0	11.4	11.5	11.8	10.5	9.8	10.7	6.7	*0.7 ± 1.0*	*27.6 ± 0.5*	*24.1 ± 0.4*
Pe	60.8	64.7	35.1	11.4	11.5	12.3	10.9	10.1	10.6	7.3	3.4	*1.8 ± 1.2*	*18.3 ± 0.4*
Pm	62.4	65.5	36.2	12.0	12.1	12.9	11.5	10.6	11.1	7.2	3.5	0.5	*0.4 ± 0.2*

Mean and standard deviation are presented as per cent. On 18S rDNA level no intraspecific variation and thus no standard deviation is given.
Ac, *Acartia clausi*, At, *A. tonsa*, Ab, *A. bifilosa*, Cham, *Centropages hamatus*, Ct, *C. typicus*, Ea, *Eurytemora affinis*, Ic, *Isias clavipes*, Ap, *Anomalocera patersoni*, Tl, *Temora longicornis*, Pp, *Paracalanus parvus*, Ch, *Calanus helgolandicus*, Pe, *Pseudocalanus elongatus* and Pm, *Pseudocalanus moultoni*.

fingerprints is further highlighted on the Bray–Curtis similarity matrix visualized as heatmap (Fig. 4d). Consequently, as found in previous studies (i.e. Feltens *et al.* 2010; Riccardi *et al.* 2012), different species, even closely related ones, can be successfully discriminated based on these species-specific proteome profiles. Regarding taxonomy on higher level, *Centropages* congeners are more similar to one another; on the contrary, this was not observed in the two *Acartia* species (Figs 4d and 5). If phylogenetic relationships can be tracked as it has been demonstrated for fruit flies (Feltens *et al.* 2010), then they must be tested on higher coverage of closely related calanoid species in future.

On intraspecific level, significant differences were observed between the different developmental stages. Nauplii and early copepodite stages (CI–V) differed from adult females and males (CVIF, CVIM) tested by ANOSIM and Mann–Whitney-test. These differences were found in *C. hamatus* (ANOSIM: $R = 0.835$, $P = 0.001$; Mann–Whitney-test: $P < 0.0001$), *P. elongatus* ($R = 0.642$, $P = 0.001$; $P < 0.0494$), *A. clausi* ($R = 0.916$, $P = 0.001$; $P < 0.0001$), *T. longicornis* ($R = 0.709$, $P = 0.001$; $P < 0.0001$), *A. patersoni* ($R = 0.506$, $P = 0.002$; $P = 0.0072$) and *C. helgolandicus* ($P < 0.0001$). Thus, quite different from species discrimination based on nucleotide sequences, intraspecific differences/variability were detected between different developmental stages, mainly between naupliar to early copepodite stages and adult females and males. This result indicates possible differences in the presence of proteins, thus stage-specific make-ups in expressed proteins, like it was demonstrated on protein spectra differences for fruit fly males and females (Feltens *et al.* 2010). Moreover, even non-morphologically identified nauplii clustered to species level, indicating the stability and power of this approach. Similar results have been also demonstrated for closely related freshwater copepods (Riccardi *et al.* 2012), showing lower differences in peak pattern between different developmental stages on intraspecific level than between species. However, the high resolution of MALDI-TOF MS for detecting stage-specific differences may find useful applications in various ecological and population studies.

No significant differences were observed between specimens of one species sampled at different seasons, years and stations. As a consequence, species clusters are not overlaid by trophic or feeding impacts/effects when analysing whole organisms, and gut content could be excluded as a possible contamination factor.

Because the number of studies on metazoan identification based on proteome profiles has recently increased (e.g. Mazzeo *et al.* 2008; Feltens *et al.* 2010; Kaufmann *et al.* 2011; Riccardi *et al.* 2012; Volta *et al.* 2012), frequent and routine usage of this method for metazoan species

Table 3 ANOSIM: differences (statistic *R* values) between the 11 species based on pairwise tests

	Ac	At	Cham	Ct	Ea	Ic	Ap	Tl	Pp	Ch
Ac										
At	**1**									
Cham	**0.986**	**1**								
Ct	**0.999**	**1**	**0.966**							
Ea	**1**	**1**	**0.987**	**1**						
Ic	**1**	*1*	**1**	*1*	*1*					
Ap	**1**	**1**	**0.997**	**1**	**1**	**1**				
Tl	**0.951**	**0.968**	**0.917**	**0.923**	**0.944**	**0.967**	**0.974**			
Pp	*0.960*	*1*	*0.939*	*1*	*1*	**1**	*1*	**0.936**		
Ch	**0.999**	**1**	**0.988**	**1**	**1**	**1**	**1**	**0.921**	*1*	
Pe	**0.966**	**0.972**	**0.953**	**0.995**	**0.985**	**0.985**	**0.991**	**0.930**	*0.982*	**0.982**

Regular and bold: $P = 0.001$, italics: $P < 0.05$, bold and grey: not significant (too low sample size).
Species abbreviations see Table 2.

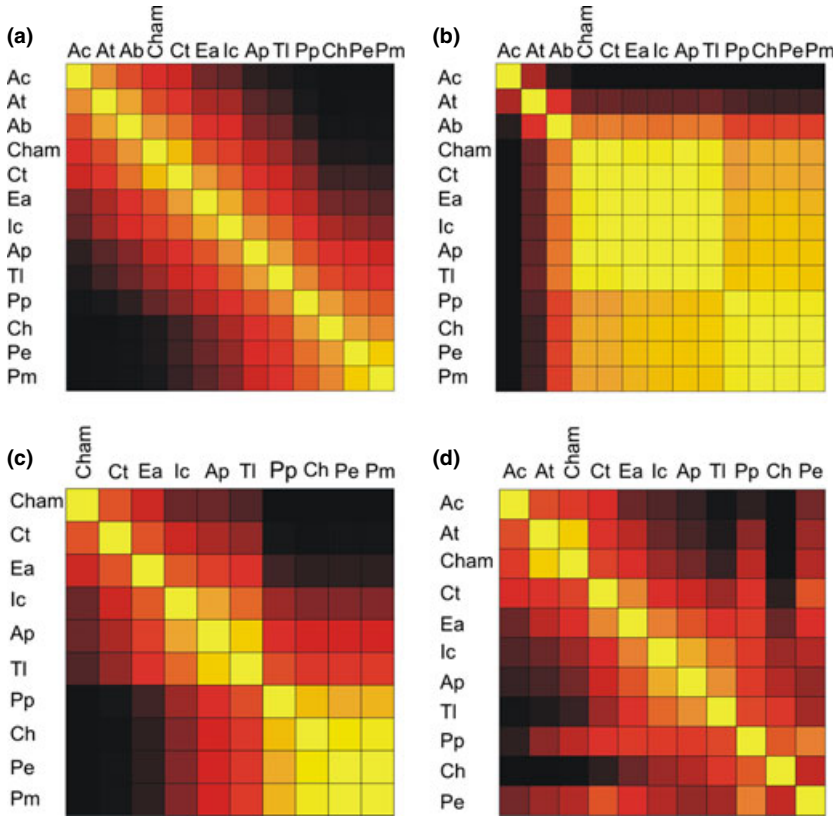

Fig. 4 Heatmaps visualizing species clusters. (a) cytochrome *c* oxidase subunit I (COI) based on mean K2P distances, (b) 18S rDNA based on GTRGAMMA, (c) 18S rDNA based on GTRGAMMA without *Acartia* species and (d) similarity matrix of proteome fingerprints (5-Da categories). Species abbreviations see Table 2. Bright colours (yellow) indicate high, dark colours (black) low similarity.

identification can be expected. In this context, the need for comparing metazoan proteome profiles for identification will increase, and thus the demand for an accessible proteome profile reference library for Metazoa (i.e. comparable to GenBank for sequence data). On this basis, MALDI-TOF MS would provide a powerful, rapid and attractive tool for molecular metazoan species identification in the same way as it is established in diagnostic bacteriology and can find its way to be integrated in taxonomic approaches.

Summarizing the results, the identification/discrimination of the 11 calanoid copepod species based on

Fig. 5 Cluster analysis of quantitative proteome fingerprints (5-Da categories) of 197 individuals of 11 calanoid species. Grey lines and bars indicate nauplii and early copepodite stages (I–V).

protein mass fingerprints was successful, irrespective of sampling origin and time, as well as developmental stage. As a consequence, this molecular profiling can be considered as a promising additional or alternative fast and accurate tool suitable for copepod species authentication. The fact that organisms or tissues fixed in ethanol can be used for typing species allows the application of MALDI-TOF MS on material, which is additionally suitable for molecular genetic studies, and thus applicable in integrated taxonomic studies.

Comparison of molecular methods

In general, molecular identification methods are objective and can be used independently of the taxonomic background of the identifier, providing a precise, digital description. In most cases, even adult organisms with missing diagnostic morphological characters (e.g. sibling and cryptic species) or early life history stages can be discriminated and thus identified within a high number of taxa. Once established for an organism group, they can be used independently from morphology. But more importantly, they can be used as a supplementary tool in integrated approaches in a broader biological–ecological context.

Comparing the investigated molecular identification methods on the basis of nucleotides and proteome fingerprints, advantages and disadvantages can be highlighted. All these three analyses were shown to be useful for a valid species identification (Figs 2–5), but with different accuracies as well as performances and realizations. DNA investigations allow unambiguous identification with sequences representing valid entities irrespectively of life history/developmental stage, origin and parts of the organism. Once DNA is extracted, it can be stored long-term and serves as a resource for various molecular genetic studies. Thus, according to the choice of the analysed gene fragment, sequences can be used for analysing haplotype diversity, phylogeographic patterns or phylogenetic relationships. Compared with COI, the analysis of 18S rDNA has the advantage of no intraspecific variability and further information on higher taxonomic level but the gap between intraspecific and interspecific variability is less pronounced (i.e. *Pseudocalanus* congeners). Due to more pronounced differences between intraspecific and interspecific pairwise genetic variation as well including less sequencing reactions for the investigated copepods, COI analysis is more suitable for species identification compared with 18S rDNA. However, disadvantages of both genetic approaches are the time-consuming and cost-intensive multiple preparation steps, including DNA extraction, PCR, purification, sequencing as well as prior methodological establishment for taxon-specific amplifications (e.g. specific primer design). Compared with this, protein mass fingerprinting analysis like MALDI-TOF MS includes less preparation steps, resulting in a rapid and cost-effective application. In

comparison with nucleotide analyses, MALDI-TOF MS approach analyses many more specimens in a shorter time period, leading to high coverage either of specimens (i.e. for population studies) or species (i.e. diversity studies). But especially for metazoans, information on the origin and nature of each analysed peptide/protein is still lacking compared with prokaryotes where half or fewer of the recorded peaks and thus proteins–peptides (3–30 kDalton in *Escherichia coli*) were suggested to be assigned to ribosomal proteins (e.g. Arnold & Reilly 1999; Ryzhov & Fenselau 2001). Nevertheless, even though the proteome profiles can be regarded as no direct reference to the analysed entities (i.e. on each analysed peptides/proteins), the protein peak pattern can be used for re-identification with the possibility to highlight typical biomarkers, as it was shown for fish species (e.g. Volta *et al*. 2012). When processing and analysing whole organisms, a disadvantage of this method can be the complete consumption of the extract, no longer available for further applications or storage. This is especially the case for small metazoans like copepods. Nevertheless, we suggest applying the MALDI-TOF MS approach where solely rapid and straightforward species identification is demanded, like identifying ambiguous species or stages from bulk samples needed for diversity studies or monitoring surveys. The fact that ethanol samples worked on all these approaches allows combined/integrated studies.

Conclusion

In total, 333 individuals of 13 calanoid copepod species were discriminated to valid and accurate species clusters using molecular tools on the basis of nucleotide sequences (COI and 18S rDNA) and proteome fingerprints. Molecular data show less variability than protein spectra, but are more time-consuming and cost-intensive. On the contrary, mass spectra are able to discriminate between some developmental stages. As a result, we suggest the application of species typing based on proteomic fingerprinting as an alternative approach for a rapid and cost-effective method. Once an accessible reference library for metazoan proteome fingerprints is established and by working with adequate software, species can be identified straightforwardly within minutes. We suggest that this method is especially useful for analysing a high number of samples or when specimens are difficult to determine/discriminate with respect to congeners or life history stages. With MALDI-TOF MS, estimation of abundances of different copepod species can reach a higher coverage especially by including early stages and populations (e.g. recruitment, thus species-specific secondary production) and thus reach a higher resolution of sample analysis.

Acknowledgements

We wish to thank the captain and the crew of FK Senckenberg, Dominik Gloe, Bettina Walter, Klas Ove Möller, Rebekka Schüller and Inga Mohrbeck for sampling support. For taxonomic support, we thank Jasmin Renz, Terue C. Kihara and Elena Markhaseva. Thank you to Viola Siegler for developing Excel-Macro for peak list data handling and to Ryan Coker for improving the English. This work was funded by the Federal Ministry of Education and Research (Grant No. 03F0499A) and the Land Niedersachsen. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

References

Aarbakke ONS, Bucklin A, Halsband C, Norrbin F (2011) Discovery of *Pseudocalanus moultoni* (Frost, 1989) in Northeast Atlantic waters based on mitochondrial COI sequence variation. *Journal of Plankton Research*, **33**, 1487–1495.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Arnold RJ, Reilly JP (1999) Observation of *Escherichia coli* ribosomal proteins and their posttranslational modifications by mass spectrometry. *Analytical Biochemistry*, **269**, 105–112.

Beaugrand G (2004) The North Sea regime shift: evidence, causes, mechanisms and consequences. *Progress in Oceanography*, **60**, 245–262.

Beaugrand G, Ibañez F, Lindley JA, Reid PC (2002) Diversity of calanoid copepods in the North Atlantic and adjacent seas: species associations and biogeography. *Marine Ecology Progress Series*, **232**, 179–195.

Blanco-Bercial L, Bradford-Grieve J, Bucklin A (2011) Molecular phylogeny of the Calanoida (Crustacea: Copepoda). *Molecular Phylogenetics and Evolution*, **59**, 103–113.

Bradford-Grieve JM, Boxshall GA, Ahyong ST, Ohtsuka S (2010) Cladistic analysis of the calanoid copepoda. *Invertebrate Systematics*, **24**, 291–321.

Bucklin A, Frost BW (2009) Morphological and molecular phylogenetic analysis of evolutionary lineages within *Clausocalanus* (Copepoda: Calanoida). *Journal of Crustacean Biology*, **29**, 111–120.

Bucklin A, Frost BW, Kocher TD (1995) Molecular systematics of six *Calanus* and three *Metridia* species (Calanoida: Copepoda). *Marine Biology*, **121**, 655–664.

Bucklin A, Bentley AM, Franzen SP (1998) Distribution and relative abundance of *Pseudocalanus moultoni* and *P. newmani* (Copepoda: Calanoida) on Georges Bank using molecular identification of sibling species. *Marine Biology*, **132**, 97–106.

Bucklin A, Astthorsson OS, Gislason A, Allen LD, Smolenack SB, Wiebe PH (2000) Population genetic variation of *Calanus finmarchicus* in Icelandic waters: preliminary evidence of genetic differences between Atlantic and Arctic populations. *ICES Journal of Marine Science*, **57**, 1592–1604.

Bucklin A, Frost BW, Bradford-Grieve J, Allen LD, Copley NJ (2003) Molecular systematic and phylogenetic assessment of 34 calanoid copepod species of the Calanidae and Clausocalanidae. *Marine Biology*, **142**, 333–343.

Bucklin A, Hopcroft R, Kosobokova K *et al*. (2010a) DNA barcoding of Arctic Ocean holozooplankton for species identification and recognition. *Deep-Sea Research II*, **57**, 40–48.

Bucklin A, Ortman BD, Jennings RM *et al*. (2010b) A "Rosetta Stone" for metazoan zooplankton: DNA barcode analysis of species diversity in the Sargasso Sea (Northwest Atlantic Ocean). *Deep-Sea Research II*, **57**, 2234–2247.

Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA barcoding of marine metazoa. *Annual Review of Marine Science*, **3**, 471–508.

Calbet A, Carlotti F, Gaudy R (2007) The feeding ecology of the copepod *Centropages typicus*. *Progress in Oceanography*, **72**, 137–150.

Castellani C, Lindley AJ, Wootton M, Lee CM, Kirby RR (2012) Morphological and genetic variation in the North Atlantic copepod, *Centropages typicus*. *Journal of the Marine Biological Association of the United Kingdom*, **92**, 99–106.

Caudill CC, Bucklin A (2004) Molecular phylogeography and evolutionary history of the estuarine copepod, *Acartia tonsa*, on the Northwest Atlantic coast. *Hydrobiologia*, **511**, 91–102.

Chen G, Hare MP (2011) Cryptic diversity and comparative phylogeography of the estuarine copepod *Acartia tonsa* on the US Atlantic coast. *Molecular Ecology*, **20**, 2425–2441.

Clarke KR, Gorley RN (2006) *PRIMER v6: User Manual/Tutorial*. PRIMER-E, Plymouth.

Dam HG, Lopes RM (2003) Omnivory in the calanoid copepod *Temora longicornis*: feeding, egg production and egg hatching rates. *Journal of Experimental Marine Biology and Ecology*, **292**, 119–137.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Edwards M, Richardson AJ (2004) Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature*, **430**, 881–884.

Eickbush TH, Eickbush DG (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics*, **175**, 477–485.

Feltens R, Görner R, Kalkhof S, Gröger-Arndt H, von Bergen M (2010) Discrimination of different species from the genus Drosophila by intact protein profiling using matrix-assisted laser desorption ionization mass spectrometry. *BMC Evolutionary Biology*, **10**, 95.

Fenselau C, Demirev PA (2001) Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrometry Reviews*, **20**, 157–171.

Folmer OM, Black W, Hoen R, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Fransz HG, Colebrook JM, Gamble JC, Krause M (1991) The zooplankton of the North Sea. *Netherlands Journal of Sea Research*, **28**, 1–52.

Frost BW (1989) A taxonomy of the marine calanoid copepod genus *Pseudocalanus*. *Canadian Journal of Zoology*, **67**, 525–551.

Goetze E (2003) Cryptic speciation on the high seas; global phylogenetics of the copepod family *Eucalanidae*. *Proceedings of the Royal Society of London. Series B*, **270**, 2321–2331.

Goetze E (2010) Species discovery in marine planktonic invertebrates through global molecular screening. *Molecular Ecology*, **19**, 952–967.

Grabbert S, Renz J, Hirche H-J, Bucklin A (2010) Species-specific PCR discrimination of species of the calanoid copepod *Pseudocalanus*, *P. acuspes* and *P. elongatus*, in the Baltic and North Seas. *Hydrobiologia*, **652**, 289–297.

Greve W, Reiners F, Nast J, Hoffmann S (2004) Helgoland Roads meso- and macrozooplankton time-series 1974 to 2004: lessons from 30 years of single spot, high frequency sampling at the only off-shore island of the North Sea. *Helgoland Marine Research*, **58**, 274–288.

Haag AM, Taylor SN, Johnston KH, Cole RB (1998) Rapid identification and speciation of *Haemophilus* bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Journal of Mass Spectrometry*, **33**, 750–756.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B*, **270**, 313–321.

Hickel W (1975) The mesozooplankton in the Wadden Sea of Sylt (North Sea). *Helgoländer Wissenschaftliche Meeresuntersuchung*, **27**, 254–262.

Hill RS, Allen LD, Bucklin A (2001) Multiplexed species-specific PCR protocol to discriminate four N. Atlantic *Calanus* species, with an mtCOI gene tree for ten *Calanus* species. *Marine Biology*, **139**, 279–287.

Hillis DM, Moritz C, Porter CA, Baker RJ (1991) Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science*, **228**, 308–311.

Holland RD, Wilkes JG, Ralli F *et al.* (1996) Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, **10**, 1227–1232.

Kaufmann CK, Ziegler D, Schaffner F, Carpenter S, Pflüger V, Mathis A (2011) Evaluation of matrix-assisted laser desorption/ionization time of flight mass spectrometry for characterization of *Culicoides nubeculosus* biting midges. *Medical and Veterinary Entomology*, **25**, 32–38.

Kiesling TL, Wilkinson E, Rabalais J, Ortner PB, McCabe MM, Fell JW (2002) Rapid identification of adult and naupliar stages of copepods using DNA hybridization methodology. *Marine Biotechnology*, **4**, 30–39.

Kimura M (1980) A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Laakmann S, Auel H, Kochzius M (2012) Evolution in the deep sea: biological traits, ecology and phylogenetics of pelagic copepods. *Molecular Phylogenetics and Evolution*, **65**, 535–546.

Lee CE (2000) Global phylography of a cryptic copepod species complex and reproductive isolation between genetically proximate "populations". *Evolution*, **54**, 2014–2027.

Lindeque PK, Harris RP, Jones MB, Smerdon GR (1999) Simple molecular method to distinguish the identity of *Calanus* species (Copepoda: Calanoida) at any developmental stage. *Marine Biology*, **133**, 91–96.

Longhurst A (1985) The structure and evolution of plankton communities. *Progress in Oceanography*, **15**, 1–35.

Machida RJ, Hashiguchi Y, Nishida M, Nishida S (2009) Zooplankton diversity analysis through single-gene sequencing of a community sample. *BMC Genomics*, **10**, 438.

Markhaseva EL, Abramova AA, Mingazov ND (2012) *Pseudocalanus acuspes* (Crustacea: Copepoda) from the White Sea. *Proceedings ZIN*, **316**, 57–70.

Mauchline J (1998) The biology of calanoid copepods. *Advances in Marine Biology*, **33**, 1–710.

Mazzeo MF, De Giulio B, Guerriero G *et al.* (2008) Fish authentication by MALDI-TOF Mass Spectrometry. *Journal of Agricultural and Food Chemistry*, **56**, 11071–11076.

Möllmann C, Kornilovs G, Fetter M, Köster FW, Hinrichsen H-H (2003) The marine copepod *Pseudocalanus elongatus* as a mediator between climate variability and fisheries in the Central Baltic Sea. *Fisheries Oceanography*, **12**, 360–368.

Radulovici AE, Archambault P, Dufresne F (2010) DNA barcodes for marine biodiversity: moving fast forward? *Diversity*, **2**, 450–472.

Raupach MJ, Mayer C, Malyutina M, Wägele J-W (2009) Multiple origins of deep-sea Asellota (Crustacea: Isopoda) from shallow waters revealed by molecular data. *Proceedings of the Royal Society of London. Series B*, **276**, 799–808.

Razouls C, de BF, Kouwenberg J, Desreumaux N (2005-2012) Diversity and Geographic Distribution of Marine Planktonic Copepods. http://copepodes.obs-banyuls.fr/en.

Riccardi N, Lucini L, Benagli C, Welker M, Wicht B, Tonolla M (2012) Potential of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) for the identification of freshwater zooplankton: a pilot study with three *Eudiaptomus* (Copepoda: Diaptomidae) species. *Journal of Plankton Research*, **34**, 484–492.

Rocha-Olivares A, Fleeger JW, Foltz DW (2001) Decoupling of molecular and morphological evolution in deep lineages of a meiobenthic harpacticoid copepod. *Molecular Biology and Evolution*, **18**, 1088–1102.

Ryzhov V, Fenselau C (2001) Characterization of the protein subset desorbed by MALDI from whole bacterial cells. *Analytical Chemistry*, **73**, 746–750.

Saiz E, Calbet A (2011) Copepod feeding in the ocean: scaling patterns, composition of their diet and the bias of estimates due to microzooplankton grazing during incubations. *Hydrobiologia*, **666**, 181–196.

Savitzky A, Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, **36**, 1627–1639.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved PCR primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Volta P, Riccardi N, Lauceri R, Tonolla M (2012) Discrimination of freshwater fish species by Matrix-Assisted Laser Desorption/Ionization-Time Of Flight Mass Spectrometry (MALDI-TOF MS): a pilot study. *Journal of Limnology*, **71**, 6.

Winkler G, Dodson JJ, Lee CE (2008) Heterogeneity within the native range: population genetic analyses of sympatric invasive and noninvasive clades of the freshwater invading copepod *Eurytemora affinis*. *Molecular Ecology*, **17**, 415–430.

Winkler G, Souissi S, Poux C, Castric V (2011) Genetic heterogeneity among *Eurytemora affinis* populations in Western Europe. *Marine Biology*, **158**, 1841–1856.

Wirth H (2009) Framework, Public version 1.2b. [http://izbifs.izbi.uni-leipzig.de/~wirth].

S.L., T.K., G.G., P.M.A and M.J.R. designed the general approach of this study. G.G. and R.E. contributed reagents and analysis tools. The analyses were performed by S.L. and R.E. S.L. analysed the data sets and wrote the manuscript. All authors contribute in editing the manuscript.

Data accessibility

DNA sequences: Genbank accession numbers JX995145-JX995324.

Supporting information are provided for multiple sequence alignments for the COI (S1) and 18S rDNA (S2) data sets as well as for mass lists (S3) and 5-Da matrix (S4) of the proteome fingerprints.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Multiple sequence alignment of mitochondrial COI sequences. COI Sequences of the investigated specimens labeled by their Accession number in GenBank are provided as a multiple alignment as (Fasta-file).

Data S2. Multiple sequence alignment of nuclear 18S rDNA sequences. 18S rDNA sequences of the investigated specimens labeled by their Accession number in GenBank are provided as a multiple alignment (Fasta-file).

Data S3. Mass lists of MALDI-TOF MS analyses for all investigated calanoid copepod specimens exported to the software Excel**.** Labels of samples (i.e. each worksheet) are explained in the first worksheet (Sample Legend) (Excel-file). Mass lists were found on the basis of following settings: centroid peak detection algorithm, 1.5 signal to noise threshold, 300 minimum/maximum number of peaks, one m/z peak width and 90% peak height using the software flexAnalysis (ver. 3.3; Bruker Daltonics).

Data S4. Processed peak tables from MALDI-TOF MS analyses. Matrix of relative intensities (5-Da bin size classes) for the investigated calanoid copepods used as template for further analyses (Excel-file).